

~~Ref~~ Sept  
UL

# Psychological Monographs

General and Applied

499  
60

Veroff Atkinson Feld and Gurin

**The Use of Thematic Apperception  
to Assess Motivation  
in a Nationwide Interview Study**

**By**

**Joseph Veroff, John W. Atkinson,  
Sheila C. Feld, and Gerald Gurin**

*Survey Research Center, University of Michigan*

Price \$1.00

74  
12



**Edited by Norman L. Munn**

**Published by the American Psychological Association, Inc.**

# Psychological Monographs: General and Applied

Combining the *Applied Psychology* with

No.

Department

Psychological Monographs general and applied

No. 499 1960

Vol. 74

no. 12

ANNE ANASTASI  
FRANK A. BEACH  
ARNOLD M. BING  
W. J. BROGDEN  
ROBERT R. BUSH  
JOHN F. DASHIE  
JAMES J. GIBSON  
D. O. HEBB  
EDNA HEIDBREDE  
FRANCIS W. IRW

Manuscripts and correspondence should be sent to the Editor. *Psychological Monographs* and other psychological journals are published in the form of presentation as journal articles; introductory and graphs, must be kept with materials which deal with

findings may be made available through the American Documentation Institute—for details of this procedure, see the *APA Publication Manual*. Preparation of manuscripts for publication as monographs should follow the procedure given in the *APA Publication Manual*. Publication in *Psychological Monographs* is free of cost to the author, except in cases where early publication is requested or author's alterations are made in galley proofs.

ARTHUR C. HOFFMAN  
Managing Editor

HELEN ORR  
Promotion Manager

Correspondence on business matters should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St., N.W., Washington 6, D.C. Address changes must arrive by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee third-class forwarding postage.

THE USE OF THEMATIC APPERCEPTION TO ASSESS  
MOTIVATION IN A NATIONWIDE INTERVIEW STUDY<sup>1</sup>

JOSEPH VEROFF, JOHN W. ATKINSON, SHEILA C. FELD, AND GERALD GURIN

*Survey Research Center, University of Michigan*

IN recent years, a number of important explorations of the use of projective measures in the survey interview setting (Campbell, 1950; Douvan & Walker, 1956; Lansing & Heyns, 1959; Maccoby & Maccoby, 1954; Sanford, 1950; Sanford & Rosenstock, 1952; Scott, 1956; Tompkins & Miner, 1958) have foreshadowed an inevitable direction of future research in social psychology: the common use in survey studies of assessment techniques developed in experimental and clinical studies of personality dynamics. This trend can be expected to promote a more meaningful integration than now exists of the societal data continually amassed in surveys of broad populations and the research findings accumulated in the clinics and laboratories of personality research. How quickly this integration will come about will depend in large measure upon the conscientiousness with which certain serious methodological problems are faced by the investigator who leaps from the laboratory or clinic to the doorstep of a startled housewife, interviewed because she happened to be selected by the magic of probability sampling. Conceptual integration of facts from field study, clinic, and experiment will be hastened if the personality assessment devices selected from the laboratory have already been investigated with some thoroughness.

The present investigation is an attempt to foster such an integration. We present here

an examination of the methodological problems faced in the first attempt to introduce thematic apperceptive measures of three motives (*n* Achievement, *n* Affiliation, and *n* Power) into a nationwide sample survey. Considerable research has been conducted to develop these measures in the experimental setting (Atkinson, 1958; McClelland, Atkinson, Clark, & Lowell, 1953). Consequently, in adapting these measures to the survey setting, we are in a position to be guided by a fund of research findings which define their validity and highlight the methodological issues to be given serious consideration.

A decade of experimental work has shown that experimentally induced motivational states influence the content of imaginative thought in ways that can be reliably coded to yield measures of important social motives—*n* Achievement, *n* Affiliation, and *n* Power. The use of these methods of content analysis in studies to assess individual differences in the strengths of these motives has produced an accumulation of factual information concerning the nature of motivational influences on behavior—at least within the population of high school and college students who have been the chief subjects of study. While results on hand attest to the validity of the approach, considerable further experimental refinement of the present measures is, without question, required. Nevertheless, several explorations using these measures in different cultural settings, i.e., in situations outside of the college laboratory (Child, Storm, & Veroff, 1958; Douvan, 1958; McClelland & Friedman, 1952; Rosen, 1956), have already suggested their promise for field research, particularly for survey studies. We realize, however, that a number of weighty metho-

<sup>1</sup> The investigations reported in this paper are based on data obtained within a national sample survey supported by the Joint Commission on Mental Illness and Health (Gurin, Project Director). Analysis of these data was supported by two grants from the National Institute of Mental Health, United States Public Health Service, Projects M2181 (Veroff, Principal Investigator) and M2280 (Gurin, Principal Investigator).

dological barriers stand in the way of using thematic apperceptive measures of motivation in a national survey, and that these barriers will only be overcome through direct study in field research.

The questions of immediate interest in this paper, therefore, have to do with methodological problems in using the thematic apperceptive method—known to require a modicum of verbal ability and known to be sensitive to the situational differences at the time of administration and known to require skill in content analysis—under the conditions of national survey interviewing. These questions are listed below.

Can a single set of pictures be selected that are sufficiently relevant to the life experiences of people in all strata of society to provide fair and unbiased measures of motivation? Can the conditions of administration, in this case the interview situation, be sufficiently controlled to approximate the standardized condition of experimental studies? Can the scoring of the content of a heterogeneous sample of imaginative stories be accomplished with the same high scoring reliability that has been attained in small-scale studies of more homogeneous populations? Can anticipated differences in verbal ability in a national sample be taken into account in a way that will yield unbiased indices of motivation? Can we estimate the importance of the effect of different interviewers on the projective measures of motivation? Each of these questions will be considered separately in the first section of this monograph.

In addition to these methodological questions, we will also consider some of the interpretive problems that are highlighted when these motivation measures are investigated within the context of a national sample. These will be discussed in the second section of this monograph.

Certainly our initial answers to these questions cannot be taken as final. We consider this paper as no more than an opening wedge, explicitly pointed to laying a firmer groundwork for further integration of insights derived from clinical, experimental, and field studies of human motivation.

## METHODOLOGICAL PROBLEMS

### *Selection of Pictures*

The particular pictures used to assess motivation by thematic apperception have important effects on the motivational content of stories (Birney, 1958; Haber & Alpert, 1958; Jacobs, 1958). Pictures will differ, depending on their content, in the average amount of motivational imagery they elicit from any group of subjects (Ss). The amount of motive-related imagery that a picture elicits from a person is influenced both by the strength of his motive and by his past experiences in settings like the one portrayed in the picture.<sup>2</sup> Variability in over-all motivation scores which can be attributed to peculiar characteristics of the pictures used rather than to the strength of motive within the Ss is a source of error we seek to avoid.

The population in the survey study to be reported in this paper is a sample of 1,619 adults (21 or over), a cross-section of Americans living in private households, selected by means of probability sampling (Kish, 1953). To construct a battery of pictures that will have similar significance for all social groups represented in this sample was the major problem of picture selection that we faced. Men and women, old people and young people, blue collar workers and white collar workers, all have had experience in different sets of life situations, often so different that the motivational significance of particular kinds of pictures is likely to vary considerably from one of these groups to another. Indeed, the sex difference is so important a factor (Angelina, 1955; Davenport, 1953; Field, 1951; Morrison, 1954; Veroff, Wilcox, & Atkinson, 1953; Vogel, 1954) that it was necessary from the very beginning to plan a different set of pictures for men and women. We hoped, in addition, to overcome other systematic biases in pictures

<sup>2</sup> Atkinson (1958, Ch. 42) has presented a theoretical scheme showing how cognitive expectations based on past experience in situations like the one portrayed in a particular picture together with the motives of the individual jointly determine the motivational content of an imaginative story.

that might be attributable to many other differences in social background, in order to allow each person a fair opportunity to express his motivation in the test as a whole. Only then could the motivation scores of different social groups be legitimately compared; this aim was a compelling interest in the initiation of our study.

How then did we go about selecting pictures for men and women separately? The list of pictures employed to study *n* Achievement, *n* Affiliation, and *n* Power (Atkinson, 1958, Appendix III) attests to the wealth of data about many pictures from assessment studies of this kind, largely on the college population. From this list we rejected pictures that were clearly oriented to the college setting (e.g., classroom scenes). Since we were interested in selecting pictures that would be relevant to three motivational contents (achievement, affiliation, power) at one time, we also rejected pictures suggesting themes that contained only one particular kind of motivational content. Initially our choices were guided only by intuition.

Before an actual pretest of some pictures in a doorstep interview, however, a more systematic evaluation of pictures was undertaken. It has been shown that ratings of pictures for strength of achievement-relatedness are correlated with the *n* Achievement scores later obtained from stories written about a picture by other groups of Ss (Birney, 1958; Haber & Alpert, 1958; Jacobs, 1958). Consequently we employed a rating procedure to ascertain empirically the cue value of many pictures. A group of college students at the University of Michigan rated 40 pictures for the salience of achievement, affiliation, power, and seven other motivational concerns. At least eight Ss rated each picture. From these ratings we tried to select pictures in which one kind of motivation was rather strongly suggested but at least one or the other kinds of motivation was weakly suggested.

Following this preliminary selection of pictures, we faced more directly the potential bias of picture cues for one social group compared to another. We hoped to

eliminate bias in the instrument attributable to pictures that allow for adequate expression of motivation by only certain groups of people. Our decision, therefore, was to select pictures portraying separately for men and women a *variety* of life situations with which, in one way or another, most people in this country have had some direct contact, viz., different types of work situations and interpersonal situations. Sampling from this variety of what appeared to us to be fairly universally-relevant situations should minimize the bias in the measuring instrument.

Combining the ratings given by the students and our *a priori* conception of diverse situations for sampling pictures, we selected a large group of pictures for pretest in a Detroit residence survey. The stories told to these pictures were then scored for *n* Achievement, *n* Affiliation, and *n* Power. The pictures selected from this group for the final test forms, one for the men and one for the women, were those producing an adequate balance of scores for the three motives. The pictures finally decided upon are listed in order of presentation<sup>3</sup> (numbers refer to the list in Appendix III, Atkinson, 1958):

#### Male Form:

1. Two men (inventors) in a shop working at a machine. (2)
2. Four men seated at a table with coffee cups. One man is writing on a sheaf of papers. (101)
3. Man (father) and children seated at breakfast table. (102)
4. Man seated at drafting board. (28)
5. Conference group. Seven men variously grouped around a conference table. (83)
6. Woman in foreground with man standing behind and to the left. (103)

#### Female Form:

1. Two women standing by a table and one woman is working with test tubes.
2. Woman (mother) seated by girl reclining in chair.
3. Group of four women. One standing, the others seated facing each other.

<sup>3</sup> These forms are available (4" x 6" pictures). They may be obtained by writing to the Survey Research Center, Ann Arbor, Michigan. Price \$1.00.

4. Woman kneeling and applying a cover to a chair.
5. Two women preparing food in the kitchen.
6. Same as 6 above.

A striking feature of these pictures is that they all provide only moderate structure to the situation portrayed. The set as a whole is much less strongly cued for one particular kind of motivation than sets of pictures used in experimental studies which have focused on only one kind of motivation. Rejected pictures were ones that were considered either under- or overstructured. The understructured pictures (e.g., a person standing alone in a doorway) did not provide sufficient cues for a story in a heterogeneous population. The overstructured pictures (e.g., a woman typing in an office) elicited stories with too little variability of response.

These two forms were administered to a group of 98 college students at the University of Michigan. The measures were administered in a latin square design controlling the order of presentation of pictures. Unlike the survey interview setting, the administration called for written stories according to procedures outlined elsewhere (Atkinson, 1958, Appendix III). We attempted to ascertain in these administrations: (a) whether possible influences on the total motivation scores of the order of presentation of pictures existed, and (b) whether the rough equivalence in sensitivity to *n* Achievement, *n* Affiliation, and *n* Power found in a sample survey held up under procedures normally employed on a sample of college students. This study showed that order of presentation of pictures did not have any systematic effect on total motive scores and that the frequencies with which achievement, affiliation, and power imagery appeared in response to each form were approximately equal in the college group. The results of these several pretests seemed to justify going ahead with the proposed set of pictures in an initial study of a national sample.

We realize, however, that the pictures finally selected undoubtedly still have differential cue values for some respondents. Negroes, for example, have to respond to

situations depicting white people. Very old and very young people have to make up stories about figures that are obviously not so old nor so young as they are. The work scenes do not include farming, surgery, ditch digging, sales, and countless other occupations, although both blue collar and white collar work situations are included. Despite these differences, our working assumptions are that the situations portrayed in the pictures finally selected are not altogether foreign to most of the people with whom we are dealing, and that, by and large, the pictures will be as suggestive of the three motivational content areas to a person of one social stratum as to a person in another. We had finally to recognize that only through intensive study of the motivational cue value of these pictures for different social groups would we sharpen our insight and discover criteria that might enable us to construct a better set and to evaluate the over-all merits of the present set of pictures. In the second section of this monograph, this problem is raised again. In that section we present motive scores for various subgroups in the national sample. Differences we may find in the scores of various subgroups can be attributed to biased selection of pictures. Whether these differences in fact reflect something more than biases in picture selection is a problem we consider in that section.

The procedures followed in administration of the pictures and in the analysis for motivational content will be described subsequently. Table 1 presents some results from the national survey that are pertinent to an evaluation of the selection of pictures. Table 1 shows the frequency of imaginative stories told in response to each picture which contained each kind of motivational content.

Both male and female forms appear more strongly cued for affiliative imagery than for achievement or power imagery. To conclude, however, that the greater affiliative response is indicative of some trend in national character would require the gratuitous assumption that each set of six pictures represents equally the populations of



TABLE 1

PERCENTAGE OF STORIES CODED FOR ACHIEVEMENT,  
AFFILIATION, AND POWER IMAGERY  
(BY PICTURE AND SEX)

Sex	Picture	Affiliation Imagery	Achievement Imagery	Power Imagery
<b>Men (N=597)</b>				
	1	3%	32%	11%
	2	16	23	14
	3	17	2	38
	4	53	13	1
	5	33	13	30
	6	33	3	22
Average Raw Score <sup>a</sup>		4.85	2.90 <sup>b</sup>	2.84
Standard Deviation		3.65	3.29	2.73
<b>Women (N=774)</b>				
	1	2%	45%	10%
	2	56	3	26
	3	44	10	6
	4	6	36	1
	5	40	30	6
	6	35	1	25
Average Raw Score <sup>a</sup>		5.31	3.97 <sup>b</sup>	1.84
Standard Deviation		3.54	3.30	2.19

<sup>a</sup> Scores based on manuals and procedures in Atkinson (1958, Chs. 12, 13, 14).

<sup>b</sup> The category, "Unrelated Imagery" was scored 0 instead of -1. Scores obtained using 0 and -1 correlate .96.

life situations in which n Achievement, n Affiliation, and n Power are normally expressed. The differences among average scores can more appropriately be used as an index of the relative sensitivity of the present pictures to each of the motives and as a guide for improving the selection of pictures in subsequent studies.

In light of Haber and Alpert's (1958) definition of a strongly cued picture as one which elicits a particular kind of motivational imagery from 75% or more of the Ss in a sample and a weakly cued picture as one which elicits imagery from between 25% and 50% of Ss, we must conclude that the two six-picture forms employed in

the present study are weakly cued for each of the three motives. None of the pictures can be considered strongly cued for any motive; and no more than four pictures in the set fall in the range Haber and Alpert designate as weakly cued for a given motive. Consequently, in the present study there should be fewer "false positives," which often occur when strongly cued pictures are used. On the other hand, the relatively low frequency of response has two important negative implications. First, scores from weakly cued pictures are necessarily less variable and therefore by themselves have less potential for generating strong relationships with other variables. Secondly, low average frequency of response to pictures is one of several considerations<sup>4</sup> that preclude the possibility of a meaningful split-half analysis of the reliability of the motive scores in this study. A split-half estimate of the reliability of the present measures is considered inappropriate because the average frequency of response is so low that even moderately motivated Ss often respond to only one or two pictures of the set.

#### *Control of the Interview Setting for the Projective Measures of Motivation*

In experimental studies using projective measures of strength of motivation, the experimenter controls the amount of time the

<sup>4</sup> Earlier systematic split-half and test-retest estimates have yielded coefficients as high as .78 (McClelland, et al., 1953) and .74 when all the pictures were strongly cued for n Achievement and .53 (Haber & Alpert, 1958, p. 661) when all the pictures fell in the 25-50% (weakly cued) range for n Achievement. In these systematic studies of reliability, care was taken to construct subsets of pictures having qualitatively as well as quantitatively similar cues. In the present study the six pictures were not deliberately selected with an eye to having two equivalent subsets of three pictures with similar content for each motive. Rather, the choice of pictures was guided by just the opposite orientation of (a) presenting a variety of nonoverlapping life situations in which (b) three different motives might be expressed. For a fuller discussion of reliability of motivation scores, as distinct from reliability of coding which is taken up at length in a subsequent section, see Atkinson (1958), Haber and Alpert (1958), McClelland (1958), and Reitman and Atkinson (1958).

*S* spends telling a story. When written forms of the test are employed, each *S* is given a booklet containing a sheet of paper for each story. On each page are four sets of questions, adapted from Murray (1953), which help to remind *S* to cover all elements of a plot:

What is happening? Who are the persons?

What had led up to this situation? That is, what has happened in the past?

What is being thought? What is wanted? By whom?

What will happen? What will be done?

Recognizing that we could not ask a great many of the potential respondents in a national survey to write out the story, we were confronted with the task of fitting a procedure that was as similar as possible to the one above into the *interview* setting. We had to face the fact that we would be dealing with respondents who differed greatly in verbal fluency, with an individual administration rather than the usual group administration,<sup>5</sup> and with interviewers having different tendencies for probing. How could we command control over these differences to provide measures of motivation that would be comparable to those obtained in well-controlled experimental settings?

In the pretest already described, we explored different wordings of the leading questions and changed them slightly to elicit the most adequate imaginative protocols. The final wording of the questions for each picture and story was:

Who are these people? (Who is this person?) What are they (is he/she) doing?

What has led up to this—what went on before?

What do they (does he/she) want—how do they (does he/she) feel?

What will happen? How will it end?

The interviewers were instructed to write in verbatim transcript the story told to each picture as responses to each of the questions. The series of questions for each story appeared on one page with a space after each for the answers given. Each interviewer had a set of 4" × 6" pictures which

he was to show to the *S*. The following excerpts from the detailed instructions given to the interviewers<sup>6</sup> reveal our decisions (a) not to make this a completely spontaneous story-telling situation but to have the interviewer ask the plot-guiding questions, and (b) not to allow the interviewers to engage in the spontaneous probing they usually engage in with open-ended questions but rather to be very specific about the conditions under which probing was permissible and what these probes should be:

#### Instructions to Interviewers

In response to direct questions about himself, a person often cannot give you an accurate picture of the things he is most concerned about. It has been found, however, that for some purposes a more accurate picture can be obtained from the kinds of stories he makes up in response to pictures. This story-telling procedure, then, is an attempt to do just that. In order to be able to compare people on the kinds of stories they tell, it is very important that each interviewer follow the same procedure. That is why we are giving you the explicit and detailed instructions that follow:

1. Read the instructions for the story-telling procedure to the respondent exactly as they are written on the questionnaire.

2. Show the *R* the picture and give him about 20 seconds to look at the picture before you ask him the questions that go with each picture. Keep the picture in front of him as he tells the story.<sup>7</sup>

3. There is one page for each story to be told. The same questions appear on each page. Ask the questions in order for each story. *Keep to the wording of the questions, and ask all of them even if it may seem repetitious in some instances.*

4. Problems that you may encounter and what to do about them:

a. If the person rejects the whole task right off the bat by saying "I can't do this" or "I just have no imagination," resort to something like: "Well, just try to make up anything you want. Remember there are no right or wrong answers."

b. If, in reply to one of the questions in the story form, *R* says "I've already told you that,"

<sup>5</sup> All interviewers were those regularly employed members of the Field Staff at the Survey Research Center. In addition to receiving written instructions, two-thirds of the interviewers received direct communications about problems arising in this interviewing procedure at regional conferences set up around the country.

<sup>7</sup> This is a departure from laboratory procedure in which a picture is shown for only 20 seconds before a story is told. This alteration was introduced to maintain rapport in an interview study.

<sup>6</sup> Lindzey and Heinemann (1955) have concluded from research studies of individual and group administered TATs that they are reasonably equivalent.



you can say, "Can you tell me some more about that?"

c. When a respondent says "I don't know" in reply to any of the questions about the pictures, again you might say, "You can make up anything you want. There are no right or wrong answers."

NOTE: We do not need detailed stories. All we are interested in are responses which answer the four questions for each story. If *R* answers the question, even if only in a couple of words, *do not probe*. Probing should come only in the cases mentioned above (a, b, c).

We would like to hold the story-telling for all six pictures down to a time limit of 20 to 22 minutes.<sup>8</sup> Most people will give you brief stories, but if a person begins to give a lot of detail, remind him before the next story that he doesn't have to give long stories.

The collection of thematic apperceptive stories was accomplished in the context of a study conducted for the Joint Commission on Mental Illness and Health (Gurin, Veroff, & Feld, 1960). The larger interview dealt with personal feelings of adjustment and ways people manage to handle their adjustment problems. Previous research has conclusively shown that the content of imaginative stories is highly susceptible to conditions immediately preceding the story-telling procedure. Indeed, the experimental validity of the measures depends upon their sensitivity to situationally-induced motivation. Therefore, to approximate a "neutral" test condition,<sup>9</sup> the projective measure appeared at the beginning of this personal interview. It was preceded only by a few questions about leisure-time activities, used to establish rapport and to avoid what might otherwise be too abrupt a beginning if the imaginative test was given first. Another survey study (Morgan, Snider, & Sobel, 1958) which included exactly the same measure of motivation but at the end of the interview (because of special rapport problems in that interview situation) found that the stories were considerably influenced by the specific nature of the material discussed in the questionnaire be-

fore the thematic apperceptive procedure was introduced.

The rationale given to respondents for introducing the novel story-telling procedure in a survey about "modern living" was:

Another thing we want to find out is what people think of situations that may come up in life. I'm going to show you some pictures of these situations and ask you to think of stories to go with them. The situations won't be clearly one thing or another—so feel free to think of any story you want to.

For example, here is the first picture. I'd like you to spend a few moments thinking of a story to go with it. To get at the story you're thinking of I'll ask you questions like: Who are these people? What do they want? and so on. Just answer with anything that comes to mind. There are no right or wrong answers.

The interviewers reported that no major difficulty arose with this testing procedure. Many of them commented that the respondents did not appear disturbed by this novel interviewing procedure and, in fact, seemed to enjoy it. The major question we must ask about the effectiveness of the procedure, however, is whether enough standardization of interviewing was established to rule out possible interviewer effects on the measures of motivation obtained from the stories. Would different interviewers testing the same population elicit stories that were similar in motivational content?<sup>10</sup> This problem is treated in a later section.

### *Coding the Stories*

*Training the Coders.* One of the most important methodological contributions of previous work on these projective measures of motivation has been the continued concern of research workers with attaining high coding reliability. Feld and Smith (1958) have evaluated the training procedures used to teach novices the methods of content analysis for the three motives and report that conscientious use of the training materials they describe enables coders to score stories for *n* Achievement, *n* Affilia-

<sup>8</sup> In experimental studies of students, 4 minutes are normally allowed for each story.

<sup>9</sup> A "neutral" test condition is one in which the experimenter makes no deliberate attempt to arouse or relax a particular kind of motivation before administration of the test.

<sup>10</sup> Birney (1956) has reported evidence of an "experimenter effect" on the level of achievement-related content in stories written by college students.

tion, and *n* Power with reliabilities acceptable for research purposes. They review interjudge reliabilities reported in 14 published studies using the method of content analysis to be employed in this analysis and find scoring reliabilities ranging from .66 to .96 with a median of .89. They also report interjudge reliabilities ranging from .73 to .92 with a median of .87 for 12 novice coders who had just finished learning how to score a set of familiar pictures using their training procedures (i.e., after about 12 hours of independent practice). This survey of the degree of objectivity in coding obtained in small-scale experimental studies provides a standard against which to evaluate the coding reliabilities attained in this first attempt with a national sample.

A team of nine coders (three for each motive) with little or no previous contact with this kind of research was trained by the method prescribed by Feld and Smith. In addition to this training procedure, scoring seminars were held, mainly to establish consensus about scoring stories derived from pictures never before used in any study. Often the stories elicited from a new picture demand specific scoring conventions not treated in scoring manuals. Feld and Smith (1958) report that coders show a small but nevertheless statistically significant decrease in reliability when coding novel material. Novel types of imagery were thoroughly discussed in the light of previous practice and the coders annotated their scoring manuals accordingly.

To insure good coding reliabilities even further, each coder was trained as an "expert" on two pictures from the male form and two from the female form, and subsequently coded only those pictures. In this way the scoring of each individual coder contributed to potentially one-third of a particular motive score for each *S*. As a result, we ruled out bias of total scores attributable to the scoring idiosyncracies of a particular coder. Further experience was given each coder on pictures for which he was finally responsible in the course of scoring 400 protocols gathered in another study at the Survey Research Center (Morgan, et al., 1958). Check-coding was done on a

sample of 240 stories from 40 *Ss* by two other coders<sup>11</sup> who had previously been checked out as "expert" in motive scoring.

Scoring reliabilities from this "warm up" check-coding, and for others reported in this paper were computed in two ways. First, we estimated the degree of agreement in noting the presence of any achievement, affiliation, or power imagery in the story—the single most important scoring decision for the coder. This agreement is expressed as a percentage (twice the number of agreements on the presence of motivational imagery divided by the sum of the total number of times motivational imagery was coded by the two coders). The second estimate of the coding reliability was the rank-order correlation between motivation scores based on all six stories coded first by the three coders and then by the check-coding experts. Although the individual percentages of agreement on "imagery" (ranging from 50% to 90%) were not consistently as high as the customary agreements found in experimental research, we felt that the total score reliabilities (.74, .76, .87) showed that these coders had attained a degree of proficiency which, augmented by further clarification of scoring problems, was sufficient to begin systematic scoring of the stories in the national sample.

*Assessing Coding Reliability of Motivation Scoring in the National Survey Sample.* The nine coders, each a specialist on four pictures (two from the male form and two from the female form) for one of the three motives, coded two stories from each of the 1,619 respondents' protocols (or 3,238 stories). The criteria for scoring or not scoring a particular response are those specified in the available manuals for *n* Achievement, *n* Affiliation, and *n* Power (Atkinson, 1958; McClelland, et al., 1953) plus specific conventions adopted for novel pictures.

Each coder's total scoring task was divided into three nearly equal sections (about 1,080 stories to be scored within each section). Furthermore, each coder

<sup>11</sup> Two of the authors—Sheila Feld and Joseph Veroff.

scored the protocols in a different order, never scoring two stories from the same S in sequence.<sup>12</sup> These three sections each represents a coding period for which check-coding procedures were done. Within each third of the coding, a set of protocols were designated as reliability checks. The particular stories that were to be coded by the "expert coder" were unknown to the coders. From these protocols we estimated the reliabilities, again both as percentage of agreement in the coding of the presence of motivational imagery in the stories, and as rank-order correlations of total scores ob-

tained from six stories. After each check-coding, differences between coder and "expert" were discussed. The three reliability checks provided an opportunity to follow the progress of coding reliability over the period of several months which it took each student coder, working part-time on an hourly basis, to complete approximately 100 hours of coding. The resulting percentages of agreement and total score reliabilities in the three different periods of coding are presented in Table 2.

The reliabilities of the coders were remarkably consistent over the three sections of coding. There are occasional shifts upward and downward, but there are no consistent trends indicating any drastic effects of coding over a prolonged period. It is our opinion that the downward trends in reliability often seemed to reflect the "expert" coder's lack of adaptation to these particular pictures. During the course of coding, certain specific conventions were adopted to establish scoring criteria for these novel

<sup>12</sup> For a given motive each S's stories were scored in such a way that two were scored by one coder during the first third of coding, two in the next third by the next coder, and two in the last by the third coder. In this way any systematic bias of coding attributable to time of coding would be randomized over the Ss. The exception to this procedure is the set of stories in the reliability checks. All of the stories of these Ss were coded during the same period of coding.

TABLE 2  
SCORING RELIABILITIES FOR THE NINE CODERS AND THE THREE MOTIVATION SCORES  
IN THE NATIONAL SAMPLE SURVEY

Motive and Coders	Pictures	Percentage of Imagery Agreement (60 Stories)			Total Score Reliabilities (30 Ss)		
		Check 1	Check 2	Check 3	Check 1	Check 2	Check 3
n Achievement					.89 <sup>b</sup>	.77	.81
Coder 1	1, 6 <sup>a</sup>	80	82	90			
Coder 2	3, 4	71	75	63			
Coder 3	2, 5	87	55	82			
n Affiliation					.91	.76	.76
Coder 1	4, 5	80	80	89			
Coder 2	2, 6	83	89	78			
Coder 3	1, 3	80	73	80			
n Power					.	.72	.73
Coder 1	1, 5	.	75	67			
Coder 2	3, 6	.	73	77			
Coder 3	2, 4	.	77	84			

<sup>a</sup> To be read, Picture 1 and Picture 6 on male form and on female form.

<sup>b</sup> Spearman rank-order correlations.

<sup>c</sup> These figures were similar to those attained in the second and third reliability checks for n Power coding. Exact figures are unavailable because the check coding sheets were misplaced and cannot be re-estimated since the coders changed their original codings in the direction of the consensus after discussion.

materials. The coders were more adapted to these changes than the expert coders, who, scoring only a sample of protocols from time to time, tended to be somewhat less consistent in decisions about "novel" material.

The extent to which this is true can be evaluated in light of a second set of reliability coefficients that were computed *after* the coder and the expert check-coder had discussed every instance of disagreement and had come to a consensus of opinion about final coding. Correlations between the coders' original scorings and this consensus of opinion were uniformly higher indicating that at times it was agreed the "expert" had been in error. The post-discussion score reliabilities for first, second, and third checks were: *n* Achievement—.91, .94, .88; *n* Affiliation—.94, .84, .81; and *n* Power—.,75, .77, .77.

Consequently, we view the reliability coefficients reported in Table 2 as conservative estimates of the objectivity of the method of analysis employed. The median percentage agreement in coding motivation imagery is 80%, about 10% less than usually attained in small-scale studies with familiar materials. All of the total score reliabilities (median = .77) fall within the range found in published experimental studies, but only two of the eight are above the median of coding reliabilities reported in smaller-scale studies (*viz.*, .89). Considering the anticipated loss in coding reliability attributable to the use of several new pictures here for the first time, we were encouraged by these results. Nevertheless we anticipate some loss in sensitivity of the measuring instruments due to some coding unreliability.

It may be mentioned in passing that the coders' awareness that there would be intermittent checks seemed to be very important in guaranteeing even coding from them. Furthermore, it gave us an opportunity to focus attention on questionable scoring criteria. As a consequence, the coders were highly aware of certain idiosyncratic coding conventions they had

adopted—some of which they eventually had to change—but some of which helped us get a better picture of how the scoring criteria apply to the stories told by this kind of broad population sample in an interview setting. These "new" conventions have been integrated into scoring manuals for this set of pictures.<sup>14</sup>

#### *Correcting Motivation Scores for Variations in Length of Protocol*

The correlation between motivation scores and length of protocol has been found negligible in studies with college students when 4 minutes are allowed for writing stories (Atkinson, 1950, p. 27). At least four studies, however, have shown significant relationships, requiring some sort of correction when somewhat different test conditions are employed. Ricciuti (1954) and Ricciuti and Sadacca (1955) found correlations ranging from .48 to .59 between *n* Achievement scores and number of words in the brief protocols of high school students when no structuring questions were asked and when only 2.5 minutes were allowed for writing stories. Walker and Atkinson (1958) found a correlation of .41 between length of protocol and fear-related motivation scores in a very heterogeneous sample of soldiers when there was less than adequate control over the actual time spent writing stories. Child, Storm, and Veroff (1958) found it necessary to correct achievement scores obtained from folk tales to remove the difference attributable to length of story. We had reason to anticipate that in this heterogeneous sample there would be gross differences in the length and adequacy of protocols that might be attributed to gross differences in verbal ability of *Ss* and/or to uncontrolled differences in the behavior of interviewers. Two separate problems were anticipated: (a) how to eliminate protocols because of inadequate responses, and (b) how to correct for anticipated correlations between

<sup>13</sup> First reliability estimate on *n* Power coding unavailable (*cf.* Footnote c, Table 2).

<sup>14</sup> Available on request from Survey Research Center (Attention: Dr. Joseph Veroff), University of Michigan, Ann Arbor, Michigan.

motivation scores and length of stories in protocols otherwise judged to be adequate for further analysis.

*Elimination of Protocols.* At the outset of scoring the protocols it was clear that some Ss gave responses which had to be considered inadequate for arriving at valid motivation scores. A few Ss tended to reject the whole undertaking, some refused to give stories to some of the pictures, some were unable or unwilling to give answers to certain questions within a story. In order to deal with this problem, the coders were instructed to keep a detailed record of "inadequate responses," i.e., not answering certain questions in a story, saying "I don't know," or "It's hard to tell," or the like. For any S there are 24 possibilities for an inadequate response—four questions on each of six stories. Table 3 shows the frequency distribution of inadequate responses in terms of the criterion for elimination of protocols that was employed.

What does an inadequate response mean for assessment of motivation? It means that the person gave no imaginative content that could possibly be scored for motivational content. The person, for whatever reason, has prevented the measuring instrument from being applied. If a person rejects the whole task (24 inadequate responses) his protocols are obviously not suitable for scoring. Even if the person rejects the task for only one picture (i.e., four inadequate responses in only one story), his protocols should not be considered, for he has not taken the complete test. Had he responded to that particular picture, his story might have been saturated with a particular kind of motivational imagery. But what of the other possibilities? Neglecting to answer one question in a few stories, neglecting to answer two questions in one story, and so on? In Table 3 we have grouped the various possibilities of inadequate responses according to whether they are judged to be omissions of minor significance, possibly serious, or serious. The judgment of the seriousness of omission was made on the basis of the extent to which each kind of omission might *potentially* affect the total motivation score. We

TABLE 3

FREQUENCY OF INADEQUATE RESPONSES TO ONE OR MORE OF THE FOUR QUESTIONS ASKED TO GUIDE THE STORY TOLD ABOUT EACH OF SIX PICTURES

Adequacy of Responses	Men (N=715)	Women (N=904)
All stories complete	491	654
Omissions Judged to Be of Minor Significance		
One story has 1 inadequate response	73	89
Two stories have 1 inadequate response	23	23
Three stories have 1 inadequate response	10	8
Omissions Judged to Be Possibly Serious		
More than 3 stories have 1 inadequate response	6	7
One story has 2 inadequate responses	35	32
Omissions Judged to Be Serious		
Two or more stories have 2 inadequate responses	14	15
One or more stories has 3 inadequate responses	25	39
One or more (but not all) stories have 4 inadequate responses	20	26
All 6 stories have 4 inadequate responses	18	11

considered this question: Had the respondent not made such omissions on his protocols, could his potential total score have been appreciably higher? Not giving a response to one question in a story does not subtract appreciably from a person's potential total score for that motive. Four such omissions might make an appreciable difference, and were therefore classified as possibly serious.

In light of our experience with practices informally employed in experimental studies we eliminated Ss judged to have possibly serious or serious omissions. Thus an S's protocols were judged to be adequate when at least half the stories were complete and no more than one question was omitted from any or all of the remaining stories.



This criterion led to the elimination of 17% of the men and 14% of the women. Two things should be kept in mind in appraising the significance of this number of rejected protocols. First, we deliberately adopted a fairly stringent criterion for elimination. Secondly, these *Ss* were eliminated not on the basis of just one set of responses to one picture, but on the basis of inadequacy of response on *any one of six pictures*. Consequently, we judge that the frequency of inadequate protocols is not out of line with the frequency of inadequate responses that would be obtained from any complex open-ended questionnaire in a nationwide survey setting if we were to eliminate *Ss* giving an inadequate response on *any one* of a set of six complex open-ended questions.

One very important question to ask about this elimination procedure is whether it has eliminated *Ss* selectively from particular social groups. The psychological variable expected to have the most important bearing on ability to answer the questions about pictures is verbal fluency. The demographic characteristic most related to this variable is education of the respondent. Table 4 shows the relationship between level of education and percentage loss of data due to

inadequate protocols. The loss of data is most severe in the least educated groups. Of those who have little or no schooling, nearly one-third of the sample of men and women are rejected from further analysis. What these data suggest, therefore, is that the measuring device as here employed in a survey setting is less appropriate for certain segments of the population—particularly among the least educated groups. This is hardly a startling result. Practically any verbal measure runs into the same difficulty. Inadequate responses to open-ended attitude questions tend to stem largely from uneducated groups; inadequate answers (response sets) on certain forced choice questionnaires also are most apparent in uneducated groups (Christie, 1958).

In the future use of these measures, therefore, one has to realize that a survey population responding to them appropriately will be biased. Any variables associated with education will have biased representation. Further examination of our data also indicates that the loss of *Ss* is heaviest from the low-income groups, the occupation groups of lower status (unskilled workers in particular), and the older age levels. All of these findings are consistent with the major finding on education, since these other groups that suffer the largest depletion in *Ss* are groups that have a heavy proportion of respondents with little education.

We suspect that a failure to produce a meaningful protocol may also reflect deficiencies in the kinds of social motivation that are being assessed in this study—particularly achievement motivation. Hence, we may be overestimating the average level of motivation in any group from which a great many *Ss* are removed because of their inadequate protocols. This additional possibility argues against using subsequent data from this analysis to estimate *absolute* norms in specific groups, but it does not vitiate the possibility of comparing social groups on average motivation scores, once we recognize that we may have biased estimates in certain groups.

*Correction for Correlation between Length of Stories in Adequate Protocols*

TABLE 4

THE RELATIONSHIP OF EDUCATION TO PERCENTAGE LOSS OF SUBJECTS DUE TO INADEQUATE PROTOCOLS

Educational Level	Men		Women	
	<i>N</i> in Sample <sup>a</sup>	Percentage Loss	<i>N</i> in Sample	Percentage Loss
No schooling; grade school (1-6 years)	107	32	125	28
Grade school (7-8 years)	138	25	160	18
High school (9-11 years)	139	12	209	16
Completed high school	169	12	266	7
Attended college	157	7	136	7

<sup>a</sup> Does not include *Ss* whose educational level was not ascertained.



TABLE 5  
NUMBER OF WORDS IN SIX STORIES FOR  
PROTOCOLS JUDGED ADEQUATE

Number of words	Men ( <i>N</i> =597)	Women ( <i>N</i> =774)
80-169	32	40
170-229	117	171
230-289	147	215
290-349	127	157
350-409	77	101
410-469	45	49
470-529	23	21
530-589	11	10
590-649	11	7
650 and above	7	3
Median	291	279

and Motivation Scores. The number of words in protocols accepted for 597 men and 774 women ranged from 80 to 823 as shown in Table 5. The median number of words was 291 for men and 279 for women, or about 46-48 words per story. This is exactly the same length obtained by Ricciuti and Sadacca (1955) from 2.5-minute stories written by high school students, but is approximately half the length usually obtained in studies with college students when 4 minutes are allotted for writing each story. Although there is no way to compare directly the time needed to write and to tell stories, one may note that the approximate time allotted for each story in the instructions given interviewers (3 to 3.5 minutes) is less than the 4 minutes normally allotted in studies of college students.

The raw *n* Achievement, *n* Affiliation, and *n* Power scores<sup>15</sup> obtained from the content analysis were correlated with length of protocol separately for men and women. The results are shown in Table 6. The results show the anticipated significant relationships between length of protocol and raw motive scores, product-moment correlations ranging between .21 and .28. Clearly

<sup>15</sup> A total score is obtained by summing the frequencies of particular kinds of motivational content appearing in each of the six stories (see Atkinson, 1958, Part II).

some correction of the raw scores is necessary to eliminate the obvious fact that a person who tells longer stories stands a greater chance of obtaining a high score based on a frequency count of particular kinds of motivational content in what he has said.

Before applying correction procedures, we simplified our considerations of the motivation scores by putting the *n* Achievement, *n* Affiliation, and *n* Power scores on the same scale within each sex. There is no legitimate basis for making direct comparisons of raw scores between men and women on a particular motive, or between the raw motivation scores within each sex. The varying means and standard deviations of raw scores, for each sex and on each motive, are partly a function of the pictures cues, as already pointed out. Putting each motive distribution on the same scale with the same means and standard deviations allows for certain comparisons. This was done by assigning percentile ranks to each motivation score (for men and women separately). When a number of persons had the same score, the median percentile rank of the interval covered was assigned to all tied scores.

Each distribution of percentile ranks was normalized. Percentile ranks were converted to normal deviate scores based on a mean of 50 and a standard deviation of 10 (*T* scores). As before each sex was considered separately.

These *T* score conversions of percentile ranks effectively remove the differences between average motivation scores for both

TABLE 6  
PRODUCT-MOMENT CORRELATIONS BETWEEN RAW  
MOTIVATION SCORES AND LENGTH OF PROTOCOL

Motive	Men ( <i>N</i> =597)	Women ( <i>N</i> =774)
<i>n</i> Achievement	.28	.25
<i>n</i> Affiliation	.20	.24
<i>n</i> Power	.21	.21

Note.—All *p* values <.0001.

men and women,<sup>16</sup> but they do not correct for the influence of length of protocol. The correlations between length of protocol and *T* scores are all within  $\pm .01$  of the original correlations between raw scores and length of protocol.

The question as to how to remove the effect of length of protocol was resolved in light of the following argument. The need for some kind of correction is obvious. A person's chance of getting a higher score on any motive is greater the longer the story he tells; the greater the amount of imagery he produces in an imaginative story, the greater the chance of it being coded as evidence of some motive. Theoretically, the best possible correction would be one based on the average of correlations between scores for many kinds of motives and the length of protocol. This average correlation would indicate the systematic relationship between length of protocol and score on any motive. Some types of motivation, e.g., *n* Achievement, might be meaningful determinants of the length of a story. A person highly motivated to achieve might want to try to tell a long and complete story in order to tell the best possible story. The correlation between *n* Achievement (content) and length of protocol might then be expected to be higher than the correlation between some other types of motivation and length of protocol. If we were to use the correlation between *n* Achievement scores and length of protocol to correct scores for *n* Achievement, it might obscure a real relationship between strength of achievement motive and length of protocol. But the average of all possible correlations would reflect the extent to which longer stories provide greater opportunity for motivational imagery of any sort to appear.

In this study, the stories have thus far been coded for only three motives. Hence the average of the three correlations provides the nearest estimate that can be made of the extent to which length of story influences opportunity to get a high score. The average of the three correlations is .22 for men and .24 for women. In each case, this average correlation was used to determine a regression line of motivation scores on length of protocol.<sup>17</sup> The average "expected" motivation score was determined for each length-of-protocol interval. The average expected score at each interval was then subtracted from the obtained average score for the whole sample on each motive separately. This difference score (obtained for each motive separately) was then used as a constant correction for all individual scores within particular intervals of length. Thus, a correction factor was added to scores that were too low because protocols were short, and a correction factor was subtracted from scores that were too high because of very lengthy protocols.

The effect of this correction was to remove the systematic expected difference in motivation scores as a function of the length of protocol. In Table 7 are presented the average *T* scores for *n* Achievement in men as a function of length of protocol before and after the correction described above was accomplished. We see in this table that the corrections remove the relationship between mean *n* Achievement scores and length of protocol. Similar effects are obtained with women on achievement motivation and with both men and women on the other motives. The over-all effects of the corrections are summarized in Table 8 which reports the existing correlations be-

<sup>16</sup> There were significant differences among the raw *n* Achievement, *n* Affiliation, and *n* Power scores both within and between sexes. These are readily explained in terms of the slight differences in degree to which the sets of pictures "suggested" each of the three motivational concerns. It seemed desirable to convert these scores immediately to the same scale, i.e., percentile rank, to avoid possible misinterpretation of the meaning of different average raw scores.

<sup>17</sup> We suspect that these three motives—*n* Achievement, *n* Affiliation, *n* Power—all might be expected to be more highly related to length of protocol given the interpersonal nature of the interview, than some other kinds of motivation (e.g., hunger). Hence we are well aware of the possibility that the correction employed may "over-correct" the scores. This seemed the lesser of two evils given the compelling a priori case that a longer protocol should increase the chance of getting a high score on any type of motivational content.

TABLE 7

EXAMPLE OF CORRECTION APPLIED TO UNCORRECTED *n* ACHIEVEMENT *T* SCORES TO REMOVE THE CORRELATION BETWEEN THESE SCORES AND LENGTH OF PROTOCOL (Males Only)

Number of Words in Protocol	<i>N</i>	Mean Uncorrected <i>n</i> Achievement Score <sup>a</sup>	Correction <sup>b</sup>	Mean Corrected <i>n</i> Achievement Score
80-169	32	45.75	3	48.75
170-229	117	47.95	2	49.95
230-289	147	50.04	1	51.04
290-349	127	49.96	0	49.96
350-409	77	52.90	-1	51.90
410-469	45	53.58	-2	51.58
470-529	23	52.96	-4	48.96
530-589	11	50.09	-5	45.09
590-649	11	55.82	-6	49.82
650 and above	7	56.43	-7	49.43

<sup>a</sup> Scores represent normal deviate equivalents using the *T* score conversion of percentile ranks, the mean is 50 and standard deviation is 10.

<sup>b</sup> Correction factor obtained from the regression line based on the average of correlations between *n* Achievement, *n* Affiliation, and *n* Power scores and length of protocol. Correction factor equals mean *n* Achievement score of the whole sample minus mean "expected" *n* Achievement score for protocols of particular length.

tween length of protocol and the final corrected motivation scores. These correlations are all low and insignificant. The final corrected scores, no longer biased by length of protocol, will henceforth be referred to as the *n* Achievement, *n* Affiliation, and *n* Power scores in all further analyses.

What effect will this correction of motivation scores for length of protocol have on subsequent analyses? The relationship of motivation scores to any variable which is in turn highly related to verbal fluency, should now be uncontaminated by the factor of verbal fluency. Where previously there might have been an artifactual relationship

between a given variable and motivation score (because both variables were related to verbal fluency), there should no longer be any relationship. Where a relationship between motivation score and a variable might previously have been obscured artifactually, there might now appear a significant relationship.

To illustrate some of the potential effects of removing the influence of length of protocol, we have examined the relationship between education and the three motivation scores before and after the correction was made. We again selected educational status as a crucial example, because it should represent that variable which is most likely to reflect differences in verbal fluency. Respondents were classified according to education as follows: grade school education only or less, some high school education but no college, and some college education.

When we first relate education to length of protocol (Table 9) we find, as expected, that there are substantial differences between the educational groupings in average length of protocols. Among men, college-educated *S*s produce longer protocols than the two other educational groups; and among women, the college group is higher

TABLE 8

PRODUCT-MOMENT CORRELATIONS BETWEEN CORRECTED MOTIVATION *T* SCORES AND LENGTH OF PROTOCOLS

Motive	Men ( <i>N</i> =597)	Women ( <i>N</i> =774)
<i>n</i> Achievement	.05	.02
<i>n</i> Affiliation	.00	.00
<i>n</i> Power	-.01	-.04

Note.—All correlations are insignificant.

TABLE 9  
MEAN LENGTH OF PROTOCOL (NUMBER OF WORDS)  
ACCORDING TO THE LEVEL OF EDUCATION  
OF SUBJECTS

Education	Men (N=593) <sup>a</sup>	Women (N=771) <sup>a</sup>
Grade school	256.4	240.8
High school	276.2	264.8 <sup>b</sup>
College	303.8 <sup>a</sup>	290.0 <sup>d</sup>

<sup>a</sup> Does not include Ss whose educational level was not ascertained.

<sup>b</sup> High school mean higher than grade school mean ( $p < .05$ ; 1-tailed test).

<sup>c</sup> College mean higher than high school or grade school mean ( $p < .05$ ; 1-tailed test).

<sup>d</sup> College mean higher than high school mean ( $p < .01$ ; 1-tailed test).

than the high school group which in turn is higher than the grade school group.

From these results we are led to anticipate that motivation scores will be related to education differently before and after correction. Table 10 presents the mean *n* Achievement, *n* Affiliation, and *n* Power scores for men in the three education groups before and after correction for differences in length of protocol. Table 11 presents a similar comparison for females.

The correction does not have any sharp differential effect on the majority of relationships between motivation scores and education before and after correction. In two instances, however, the correction does make a difference. Among men, there is no significant difference between the *n* Achievement scores of high school and college groups after correction where there was a significant difference before correction. Among women, the significant difference between the grade school group and the other groups in *n* Power scores disappears after correction. The correction seems to have eliminated a possibly artifactual difference in these two instances.

The remaining significant differences in motivation score attributable to educational status are consistent before and after correction. Evidently these are relationships between motivation and education which are independent of verbal fluency. We can,

at least, feel confident that the differences obtained after correction are not attributable to differences in verbal fluency.

### *The Effect of Interviewers on Motivation Scores*

Although we made a serious effort to standardize the interviewing procedure, we realize that having a group of 159 different interviewers (27 men and 132 women) collecting the protocols must necessarily introduce some variability into the measures that has hitherto been minimal when all protocols are collected at one time in a group administration. An estimate of the extent of this interviewer variability will help us to gain some perspective concerning how far the national survey interview departs from standardized administration procedures.

To establish such an estimate in survey data is a difficult problem. A valid test of

TABLE 10  
MEAN *n* ACHIEVEMENT, *n* AFFILIATION, AND  
*n* POWER *T* SCORES OF MALES BEFORE AND AFTER  
CORRECTION FOR VERBAL FLUENCY

Motive and Education	Mean Score <sup>a</sup>	
	Before Correction	After Correction
<i>n</i> Achievement		
Grade school	48.76	49.90
High school	49.49	50.23
College	51.95 <sup>b</sup>	51.92 <sup>a</sup>
<i>n</i> Affiliation		
Grade school	49.21	49.88
High school	49.97	50.22
College	51.27	51.27
<i>n</i> Power		
Grade school	50.97	51.38
High school	49.58	49.94
College	50.23	50.09

<sup>a</sup> Does not include Ss whose educational level was not ascertained. Grade school *N*=176, high school *N*=271, college *N*=146.

<sup>b</sup> College mean score significantly higher than high school and grade school means ( $p < .05$ ).

<sup>c</sup> College mean score significantly higher than grade school mean score ( $p < .05$ ).

TABLE 11

MEAN N ACHIEVEMENT, N AFFILIATION, AND  
N POWER T SCORES OF FEMALES BEFORE AND AFTER  
CORRECTION FOR VERBAL FLUENCY

Motive and Education	Mean Score <sup>a</sup>	
	Before Correction	After Correction
n Achievement		
Grade school	48.50 <sup>b</sup>	49.46 <sup>b</sup>
High school	50.41	51.18
College	51.49	51.67
n Affiliation		
Grade school	47.82 <sup>b</sup>	47.86 <sup>b</sup>
High school	50.82	50.33
College	51.57	50.43
n Power		
Grade school	48.77 <sup>b</sup>	49.89
High school	50.50	51.09
College	51.56	51.47

<sup>a</sup> Does not include Ss whose educational level was not ascertained. Grade school  $N=226$ , high school  $N=428$ , college  $N=127$ .

<sup>b</sup> Grade school mean score significantly lower than both high school and college mean score ( $p < .05$ ).

the effect of different interviewers requires random assignment from the population of respondents to the interviewers. This is decidedly not the case with the Survey Research Center's field operation. All interviewers are assigned respondents within their general geographical locality, and sometimes more specific locations are assigned particular interviewers. Especially in a metropolitan area, there are gross differences between the sample of respondents interviewed by one interviewer contrasted with another: some interviewers specialize in low income, foreign speaking, or Negro areas. While the selection of respondents is completely determined at the central office of the Survey Research Center by probability sampling techniques, these respondents are not assigned to interviewers at random. To the degree that the assignment is not random, there may be systematic differences in responses obtained by various interviewers that stem from differences in the composition of the samples interviewed,

and not from differences attributable to the interviewers themselves. In spite of this danger of finding effects that are really attributable to respondents, we wanted to arrive at some estimate of interviewer effect. We therefore designed an analysis of "possible" interviewer effects which would be least subject to criticism that we had grossly confounded interviewers and characteristics of respondents. Our analysis undoubtedly still can be criticized on this score. But in any case, a violation of the assumption that the respondents of each interviewer represent a random sample of the population should in most cases contribute to an overestimation of interviewer effects. As a result the estimates to be described can probably be viewed as *maximal* estimates of interviewer differences.

We have examined possible interviewer effects in a limited sample of the total population of interviewers. The limitations introduced to approximate the assumption of equivalence among respondents for each interviewer were:

1. Interviewers to be considered should be white females. We eliminated from analysis all Negro and male interviewers because both are likely to be assigned to very specific populations, Negro interviewers to Negro residences and male interviewers to poorer residences.

2. Interviewers to be considered should be doing their work in nonmetropolitan sampling points. The primary sampling unit (PSU) in this survey generally corresponds to counties in states. Some of these PSUs are metropolitan. As already stated, in metropolitan areas, there is a greater tendency for the interviewers to be selectively assigned to various subpopulations. We avoid the grossest errors in meeting the assumption of random assignment of respondents by not considering metropolitan PSUs.

3. Interviewers to be considered should have interviewed at least four male respondents and/or four female respondents who gave story protocols judged adequate for motivational analysis. A study of interviewers who interviewed fewer than four Ss would provide too unreliable an estimate of interviewer effect.

4. Interviewers to be considered should be paired with a qualifying interviewer (not eliminated by Criteria 1, 2, and 3 above) from the same PSU. We imposed this limitation to allow for interviewer differences *within* PSUs. Our failure to meet the assumptions of random selection of respondents among interviewers should be somewhat compensated by the fact that we are dealing with

a relatively homogeneous (nonmetropolitan) population of respondents in which possible interviewer effects will be examined.

Given these limitations, we were able to find 8 pairs of interviewers for male Ss and 15 pairs of interviewers for female Ss. Each pair represents a different PSU. We designed an analysis of variance of motivation scores broken down into the following components: variance attributable to Ss (or individual differences), variance attributable to interviewers, and variance attributable to PSUs. The procedure<sup>18</sup> for the analysis is described by Anderson and Bancroft (1952). The component variance attributable to interviewers is segregated from the *S* variance and its percentage contribution to *S* variability can be estimated. Separate analyses of variance were conducted for male and female Ss for each of the three motives. In these analyses, we estimate the significance of the variance assignable to

interviewer effects and, in addition, we segregate the component variance for interviewers and examine the extent of its effect on score variability. Table 12 presents these analyses for men and Table 13 for women.

Tables 12 and 13 show that a "possible" interviewer effect on motivation scores is statistically significant in three instances: *n* Achievement scores for women, and *n* Power scores for both men and women. Furthermore, when component variance for interviewers is specifically segregated, its effect on score variability is estimated as an intraclass correlation from which we can derive the proportion of variance in motivation scores possibly contributed by interviewer differences (Snedecor, 1946, p. 243). These were negligible for men except for *n* Power where the percentage increase in variance was .29. Among women, the percentage increase was negligible for *n* Affiliation but .18 for *n* Achievement and *n* Power. Considering these estimates as maximal estimates, we feel safe in concluding that we have achieved a fair degree of

<sup>18</sup> We are indebted to Leslie Kish for his suggestions with procedure.

TABLE 12  
ANALYSES OF VARIANCE OF MALES' MOTIVATION SCORES  
(Ss within Interviewers within PSUs)<sup>a</sup>

Source of Variation	<i>df</i>	<i>MS</i>	<i>F</i>	Component Variance for Interviewers	Intraclass Correlation <sup>b</sup>
<i>n</i> Achievement					
PSU	7	96.77			
Interviewer within PSU	8	97.78	1.38	6.81	.09
Ss within interviewer	48	70.42			
Total	63				
<i>n</i> Affiliation					
PSU	7	186.00			
Interviewer within PSU	8	135.13	1.33	8.47	.08
Ss within interviewer	48	101.27			
Total	63				
<i>n</i> Power					
PSU	7	220.00			
Interviewer within PSU	8	170.13	2.65*	26.51	.29
Ss within interviewer	48	64.06			
Total	63				

<sup>a</sup> Limited to cases of white, female interviewers of nonmetropolitan PSUs who interviewed at least four Ss giving adequate measures.

<sup>b</sup> Ratio of interviewer component variance to *S* variance plus interviewer component variance.

\*  $p < .05$ .



TABLE 13  
ANALYSES OF VARIANCE OF FEMALES' MOTIVATION SCORES  
(Ss within Interviewers within PSUs)<sup>a</sup>

Source of Variation	df	MS	F	Component Variance for Interviewers	Intraclass Correlation <sup>b</sup>
n Achievement					
PSU	14	118.00			
Interviewer within PSU	15	174.40	1.90*	20.70	.18
Ss within interviewer	90	91.24			
Total	119				
n Affiliation					
PSU	14	90.14			
Interviewer within PSU	15	72.13	.88	-2.33	-.03
Ss within interviewer	90	81.44			
Total	119				
n Power					
PSU	14	86.14			
Interviewer within PSU	15	107.27	1.88*	12.61	.18
Ss within interviewer	90	56.84			
Total	119				

<sup>a</sup> Limited to cases of white, female interviewers of nonmetropolitan PSUs who interviewed at least four Ss giving adequate measures.

<sup>b</sup> Ratio of interviewer component variance to S variance plus interviewer component variance.

\*  $p < .05$ .

standardization in interviewing techniques for these measures of motivation. The three significant interviewer effects will be examined in later work to explore their potential psychological significance. For the present we must recognize that we have some evidence of possible influence of interviewers on some motivation scores but apparently the influence is not substantial enough to warrant a great deal of concern.

One reason these maximal estimates of interviewer effects on motivation scores are low is that the motivation scores have been corrected for length of protocol. Similar analyses of variance were run using number of words in the protocol as the dependent variable. Table 14 shows that for both men and women respondents there is a substantial interviewer effect on length of protocol. The effect is especially strong among male Ss. The correction of motivation scores to eliminate the effect of differences in length of protocol has removed most of the possible interviewer effect on these scores.

#### *Over-all Evaluation of Methodological Problems Encountered*

These, then, are the methodological problems encountered in the survey setting and our evaluations of the extent to which they remain as problems or as sources of error in subsequent uses of the thematic apperceptive measures of motivation. One problem was solved—the bias of motivation scores attributable to verbal fluency. One problem appears very difficult to solve—the inadequate protocols produced by uneducated persons. A third problem should be resolved by further empirical analysis—possible bias in the selection of pictures. The two remaining problems—lower coding reliability than in experimental studies and possible bias in motivation scores attributable to interviewers—are ones that lower the reliability of measurement in subsequent analyses using the motivation scores. But both of these sources of error appear to be ones that certainly can be greatly reduced in subsequent studies. Experimental in-

TABLE 14  
ANALYSES OF VARIANCE OF MALES' AND FEMALES' LENGTH OF PROTOCOLS  
(*Ss* within Interviewers within PSUs)<sup>a</sup>

Source of Variation	<i>df</i>	<i>MS</i>	<i>F</i>	Component Variance for Interviewers	Intraclass Correlation <sup>b</sup>
For Males					
PSU	7	13.91			
Interviewer within PSU	8	12.92	12.99**	2.96	.73
<i>Ss</i> within interviewer	48	1.06			
Total	63				
For Females					
PSU	14	9.94			
Interviewer within PSU	15	3.50	2.00*	.67	.28
<i>Ss</i> within interviewer	90	1.75			
Total	119				

<sup>a</sup> Limited to cases of white, female interviewers in nonmetropolitan PSUs who interviewed at least four *Ss* giving adequate measures.

<sup>b</sup> Ratio of interviewer component variance to *S* variance plus interviewer component variance.

\*  $p < .05$ .

\*\*  $p < .001$ .

vestigations can be conducted to clarify the difficult scoring questions that arise with new pictures, as with those employed here for the first time, and to reduce the error in coding that is a normal consequence of trying to fit general scoring criteria to imagery produced by novel stimuli. Similarly, the influence of interviewers, particularly on length of protocols, should be greatly reduced by even more thorough briefing of interviewers. We can be guided by the results of this study concerning the conditions of administration, and, specifically, the desired length of stories.

#### INTERPRETATION OF MOTIVATION SCORES

If methodological problems of the sort we have discussed in the previous section are seriously considered, the use of thematic apperceptive measures of motivation, and other sensitive techniques of personality assessment, can be expected to begin to contribute to the study of important substantive problems in survey research. Yet even after methodological precautions are taken, there remains an important question: What conceptual meaning do we attribute to the motivation scores? This is the issue now to be considered.

The problem of interpretation of motivation scores obtained from thematic apperception is not a new one (Atkinson, 1958; McClelland, et al., 1953); but the broader setting of a national survey study exaggerates the difficulties that are normally encountered even in fairly well-controlled experimental studies. Interpretive problems are particularly highlighted, as we will attempt to point up in this section, when we examine the relationships between demographic characteristics and motivation scores. The national survey which enables us to compare the motivation scores of the different subgroups of the larger population also serves to accent theoretical and interpretive problems that are often glossed over in laboratory research.

Three factors are thought to influence the frequency of motivational imagery produced in stories elicited by pictures (taken as the measure of the strength of a particular kind of motivation): (a) individual differences in strength of motive, a relatively general and enduring disposition within the person to strive for certain kinds of goal satisfaction, affect motivational imagery; (b) cues in the pictures that are used to elicit imaginative thought may also heighten or dampen

motivation for particular goals to the extent that the situation portrayed is relevant to the life experience of the respondent; and, (c) the immediate life situation of the person may arouse or dampen motivation for particular goals. The latter influence may be some factor in the immediate test situation of the respondent, or some more general but transient influence on his life situation, or some more general and longer lasting influence on his life situation (e.g., constant achievement pressure of the job) at the time the measure is administered. Both pictures and a person's life situation provide cues that define expectations concerning what kinds of goals are relevant and reachable, and how attractive it would be to reach the goals.

In previous sections we have discussed the steps taken to minimize or control the influence of some situational and picture effects in order that motivation scores obtained in this survey could be interpreted with some confidence as manifestations of individual differences in strength of motives. Many of these steps (i.e., control of the interview setting, attention to the type of pictures employed, corrections for differences in length of protocols) do, undoubtedly, enhance the possibility of viewing the obtained scores as measures of differences in personality. Nevertheless, we believe this assumption should still be considered tentative.

While we attempted to remove biasing effects of pictures, we have no independent evidence that we have in fact removed all possible bias. For instance, can we assume that the pictures employed, which portray relatively young men and women in common life situations, will have the same meaning for young, middle aged, and aged respondents? Furthermore, the standardization of the interview situation might control the immediate situational influences at the time the test was administered but obviously could have no effect on life conditions affecting the person at the time the survey was conducted. Can we assume, in advance, that the active young businessman is being tested under situational circumstances that are comparable to those of an

older man enjoying retirement from business and the stress of day-to-day activities? Or must we conceive of the possibility that these gross differences in the total life situation in which the administration of our test of motivation is embedded will produce effects comparable to those produced by experimental manipulation of situational influences in studies of motivation in college students?

Therefore when our results show that a particular kind of motivation is stronger in one social group than another, our task is not finished; rather, the difficult task of interpretation is now thrust upon us. We may, in light of earlier research, assume that the combined effect of picture cues, situational pressures, and dispositional differences is greater in one group than in another. We have located a difference in motivation which now requires some kind of theoretical interpretation. The inference that the result means the two groups differ in strength of a general and enduring disposition of personality requires the assumption, and supportive argument, that the effects of picture cues may be considered equivalent in the two groups and that the situational pressures for expression of the motive, whether transient or relatively enduring, are also equivalent. Often the arguments in support of one or another assumption will appear thoroughly convincing. For example, the assumption has generally been made in laboratory experiments with college students that in view of the relative homogeneity of both their present and past life conditions, differences in motivation scores could be assumed to reflect dispositional differences. Just as often, however, we shall find ourselves with a problem of interpretation which cannot be settled within the framework of the present results. In these cases, the identification of important motivational differences and the ambiguity in the interpretation will define an important problem for future research.

It can be argued that certain life situations can have lasting effects on a person's disposition so that "situational" differences become "personality" differences. Little is known concerning how permanent are the

effects of certain life experiences, particularly ones that occur later in life. The assumption of the primacy of early experiences in the generation of personality dispositions has been widely accepted. Facing frankly the paucity of evidence concerning dispositional versus situational factors in motivation, and their interactions, we can see that any one of the following arguments might be pursued in the interpretation of obtained differences between certain social groups.

1. The obtained difference is attributable to differences in enduring personality dispositions acquired early in life.

2. The obtained difference is the consequence of a change in personality disposition induced by important situational factors later in life.

3. The obtained difference is a consequence of exposure to differential temporary situational pressures affecting the level of aroused motivation regardless of equivalence in underlying personality disposition.

Thus two problems remain that must be faced in confronting the substantive results obtained with the thematic apperception instrument: assessments of the bias of picture cues and the influence of the *general* life situation on the motivation scores. These problems will be illustrated in the discussion of substantive results that follow.

In Tables 15 to 22 are presented some of the obtained differences in *n* Achievement, *n* Affiliation, and *n* Power scores when certain social groups in American society are compared. The complete interpretation of these and many other results is reserved for subsequent papers. The issue of alternative interpretations can be dealt with thoroughly only when appropriate demographic controls are introduced and the whole pattern of concurrent findings is considered. These tables are presented in the present context for two reasons: first, to illustrate concretely the issues which arise when we are finally confronted with the task of arriving at a valid interpretation of the motivation scores; and secondly, to provide researchers who are interested in particular social groups some descriptive

information concerning potential differences in motivation among these groups. Under no circumstances should these data be considered unequivocal norms about strength of motives in particular social groups until the issue of the extent to which obtained differences are situational, dispositional, or artifactual is settled. Furthermore, under no circumstances should all of the differences presented in these tables be considered reliably different from one another. We have highlighted some that illustrate the problem of interpreting motive scores, but we have reserved the problem of statistical significance for the substantive appraisal of results in future articles. For readers who would like some bearings on the approximate differences in percentages required for a statistically significant comparison between two independent groups, we have included in the Appendix a table of differences required for significance (.05 level) with different sample sizes (*N*s). This table is based on approximate sampling errors of differences in percentages. Some of the analyses of the differences in percentages in the following tables require statistical treatments other than what is involved in using the table in the Appendix (correlation, matched comparisons, one-tailed tests of significance). This Appendix table is presented merely to provide approximate estimates of the statistical significance of obtained differences.

In all of the tables, *S*s have been classified as "High" or "Low" in motivation score by dividing the distribution of corrected *T* scores as near to the median of the distribution for the whole sample as was possible. The tables indicate what percentage of particular subgroups are high (above the median) in the national sample. Direct cross comparisons between men and women are not justified since different test forms were employed for the two sexes and the distributions of scores for men and women were considered separately in designating high and low scores.

#### Education

In an earlier section we examined the correction for differences in length of pro-

tocol as it affected the relationship between level of education attained by the *Ss* and mean motivation scores. In Table 15 the same result is presented in terms of the percentage of scores in each of three educational levels that are above the median for the whole sample for each motive. The percentage of college educated *Ss*, whether men or women, with high *n* Achievement scores is substantially higher than the percentage of *Ss* who have had only a grade school education. Among women, the same result appears in connection with affiliation scores and power scores.

Now how would one interpret these differences? Are the picture cues biased in favor of the college population? Has the life experience of a college education affected the achievement motive of college women? Or were college women high in achievement motivation from an early age, and thereby managed to get to college? We cannot resolve these different views. Later

evidence, however, may bear on which interpretation is more appropriate.

### Occupation

In Table 16 are presented the percentage of men presently employed on a full-time basis in various occupations who obtained high scores for each of the three motives. Achievement motivation scores are much more frequently high among persons in higher status occupations than among men in lower status occupations.

No systematic relationship between *n* Affiliation or *n* Power scores and occupational status is clearly apparent, though certain possibly meaningful differences in the different occupations may be discerned. Whereas 51% of managers and proprietors score high in power motivation, only 43% of professionals and 40% of clerical workers, the remainder of the white collar occupation category, score high in power motivation. Should the difference here be considered a manifestation of differences in basic personality disposition which has led one group of men to positions of supervision and another not, or should the difference in motivation scores be considered a reflection of the difference in motivational

TABLE 15

EDUCATIONAL LEVEL RELATED TO *N* ACHIEVEMENT,  
*N* AFFILIATION, AND *N* POWER

Motive <sup>a</sup> and Education	Percentage <sup>b</sup> of High Scores <sup>c</sup>	
	Men	Women
<i>n</i> Achievement		
Grade school	48	44
High school	49	49
College	62	52
<i>n</i> Affiliation		
Grade school	47	43
High school	47	53
College	51	52
<i>n</i> Power		
Grade school	55	42
High school	47	51
College	49	54

<sup>a</sup> Here and in all subsequent tables, the motive measures used are the *T* scores corrected for verbal fluency.

<sup>b</sup> Does not include *Ss* whose educational level was not ascertained. Grade school *Ns*, Men = 176, Women = 222; high school *Ns*, Men = 271, Women = 422; college *Ns*, Men = 146, Women = 127.

<sup>c</sup> Here and in all subsequent tables, high scores refer to the scores on each motive measure that are above the median for the total sample of men or women.

TABLE 16

OCCUPATIONAL LEVEL RELATED TO *N* ACHIEVEMENT,  
*N* AFFILIATION, AND *N* POWER  
(MEN EMPLOYED FULL-TIME ONLY)

Occupation of Respondents	<i>N</i> <sup>a</sup>	Percentage of High Scores		
		<i>n</i> Ach	<i>n</i> Aff	<i>n</i> Pow
Professionals	67	60	54	43
Managers and proprietors	70	59	56	51
Clerical workers	30	57	47	40
Sales workers	34	59	47	47
Skilled workers	120	50	55	50
Semiskilled workers	88	52	41	53
Unskilled workers	38	45	37	50
Farmers	39	44	31	51

<sup>a</sup> Does not include *Ss* whose occupational level was not ascertained or *Ss* whose occupations were not codable into these categories.

influences contained in the day-to-day life situation of managers, as a group, versus those classified as professionals? Or, could the result conceivably be attributed to some biasing in the content of the picture cues favoring one group and not another with respect to the suggestion of power-related imagery? Again, the matter of interpretation must await the extensive analysis of these and related findings that will be the substance of future papers based on this investigation.

### Income

Table 17 presents the relationships between motivation scores and family income. The irregular relationship between income and *n* Achievement scores in men clearly suggests that other demographic factors

must be controlled in an effort to clarify the result. However, oddly, in the case of women, there is a positive relationship between achievement motivation scores and family income.

### Age

Perhaps the clearest illustration of the problem of interpretation of scores as effects of personality disposition versus situation effects is apparent in consideration of age trends, particularly in the achievement motivation scores of men (Table 18).

There are some notable differences between the age groups in achievement motivation (for both men and women). For example, the highest achievement motivation is found among the youngest men, and the next highest among a middle-aged

TABLE 17

FAMILY INCOME RELATED TO *n* ACHIEVEMENT,  
*n* AFFILIATION, AND *n* POWER

Motive and Family Income	<i>N</i> <sup>a</sup>	Percentage of High Scores		<i>N</i> <sup>a</sup>
		Men	Women	
<i>n</i> Achievement				
Under \$1,999	66	42	42	122
\$2,000-3,999	114	54	42	187
\$4,000-4,999	100	47	48	125
\$5,000-6,999	158	56	50	170
\$7,000-9,999	89	48	57	96
\$10,000 and above	60	57	52	52
<i>n</i> Affiliation				
Under \$1,999		42	47	
\$2,000-3,999		54	47	
\$4,000-4,999		41	51	
\$5,000-6,999		49	55	
\$7,000-9,999		52	50	
\$10,000 and above		48	52	
<i>n</i> Power				
Under \$1,999		58	43	
\$2,000-3,999		49	46	
\$4,000-4,999		47	48	
\$5,000-6,999		48	56	
\$7,000-9,999		47	57	
\$10,000 and above		55	40	

<sup>a</sup> Does not include *Ss* whose family income was not ascertained.

TABLE 18

AGE RELATED TO *n* ACHIEVEMENT, *n* AFFILIATION,  
AND *n* POWER

Motive and Age	<i>N</i> <sup>a</sup>	Percentage of High Scores		<i>N</i> <sup>a</sup>
		Men	Women	
<i>n</i> Achievement				
21-24	32	66	47	53
25-34	165	50	50	205
35-44	119	57	50	184
45-54	123	54	45	152
55-64	82	42	52	98
65 plus	76	47	35	77
<i>n</i> Affiliation				
21-24		47	53	
25-34		48	53	
35-44		44	55	
45-54		49	53	
55-64		53	36	
65 plus		49	38	
<i>n</i> Power				
21-24		44	55	
25-34		45	52	
35-44		52	47	
45-54		51	49	
55-64		49	45	
65 plus		54	47	

<sup>a</sup> Does not include *Ss* whose age was not ascertained.



group (35-44 years), while the lowest achievement motivation scores appear in the older groups. A number of alternative interpretations of these differences can be offered. On the one hand, these differences may reflect some important generational differences in the strength of achievement motive dispositions. On the other hand, the stage of a man's life cycle may have a great deal to do with his current aroused achievement motivation—the eagerness of youth starting out in a career and the middle-aged pressure to "achieve or else" may both lead to heightened achievement motivation in contrast to the pressures of old age that may detract from achievement arousal. From other results we have obtained thus far, relating achievement motivation to other variables (controlling for age), we do find that the age groups vary considerably in what achievement motivation signifies. One can at the present only speculate about a life cycle or a generation difference influencing these results, and perhaps both sets of factors are crucial.

The age differences obtained may reflect picture biases as well; the possibility of systematic age differences in the connotation of the pictures cannot be excluded. Very old people, for example, might have special difficulty in generalizing from the pictures presented to their own life situations since the figures in the pictures are not clearly old in appearance.

As a result of these obtained age differences in motivation scores and our speculations about their meaning, we have considered age as a critical variable for investigations of other substantive relationships to be reported in future papers.

### Race

One of the most apparent instances where subgroup differences in motivation scores may reflect the operation of picture bias is the comparison of scores for white and Negro respondents (Table 19). The pictures presented to all respondents for the most part clearly portrayed white persons. What meaning does this have for a Negro person? It has been found that the socio-

economic position of the Negro can play as important a role in determination of some of his responses as does his race (Korchin, Mitchell, & Meltzoff, 1950) and that there is some tendency for Negroes to be less guarded in their responses to pictures of white persons (Cook, 1953). Other research has also explored this and related problems, but no unassailable conclusion has been drawn about the relative merits of using pictures of whites or Negroes with Negro respondents (Light, 1955; Riess, Schwartz, & Cottingham, 1950; Schwartz, Riess, & Cottingham, 1951).

Comparisons between white and Negro respondents are also limited by the immediate test situation. Some Negroes were interviewed by white interviewers, although some of them (especially those in the South) were interviewed by Negro interviewers. Previous research dictates that we make note of this difference, although its implications are not entirely clear. Schwartz, et al. (1951) find that Negroes express more ideas to a TAT if the interviewer is white.

Therefore, we do not have firm empirical ground for interpreting the meaning of the

TABLE 19  
RACE RELATED TO N ACHIEVEMENT, N AFFILIATION,  
AND N POWER

Motive and Race	Percentage of High Scores <sup>a</sup>	
	Men	Women
n Achievement		
White	51	49
Negro	48	40
n Affiliation		
White	48	51
Negro	45	50
n Power		
White	48	49
Negro	61	50

<sup>a</sup> Does not include Ss whose race was not ascertained or Ss whose race was not codable into these categories. White Ns, Men = 538, Women = 688; Negro Ns, Men = 31, Women = 60.

differences found in Table 19. At the present we can only note that white women seem to be higher in achievement motivation scores, and Negro men are higher than white men in power motivation. The other comparisons show minimal differences. We can ask would we have obtained other differences had we used pictures of Negroes? Of course we cannot answer this question in the present study. Where we will use the scores of Negroes, we will treat them as if there were no picture biases. Later results may lead us to reconsider the problem.

### Place of Residence

In Table 20 we present the distributions of motivation scores according to the place of residence—whether in a metropolitan area or their suburbs, a small city, a small town, or a rural area. These findings

TABLE 20

CURRENT PLACE OF RESIDENCE RELATED TO  
N ACHIEVEMENT, N AFFILIATION, AND N POWER

Motive and Place of Residence <sup>a</sup>	N	Percentage of High Scores		N
		Men	Women	
<b>n Achievement</b>				
Metropolitan areas	86	48	45	104
Suburbs	87	52	42	97
Small cities	92	51	43	118
Small towns	180	56	49	152
Rural areas	225	53	49	230
<b>n Affiliation</b>				
Metropolitan areas		42	58	
Suburbs		44	49	
Small cities		48	49	
Small towns		48	54	
Rural areas		51	46	
<b>n Power</b>				
Metropolitan areas		55	57	
Suburbs		56	45	
Small cities		46	46	
Small towns		44	51	
Rural areas		50	48	

<sup>a</sup> Metropolitan areas and their suburbs are defined by the United States Census Bureau Classification: small cities have populations of over 50,000, small towns populations of under 50,000, and rural areas are open farm country not located in a standard metropolitan area.

present some differences in scores that are not completely independent of status considerations noted in other comparisons—income, occupation, or education differences. Nevertheless there are noteworthy results: For both men and women there is greater concentration of high achievement motivation scores in small towns and rural areas than in larger communities. High affiliation motivation seems to be more prevalent among men from smaller communities, but highest in women living in metropolitan areas. Power motivation seems to be relatively high in metropolitan areas.

These differences may have important implications for the thesis that the type of community one lives in can determine the development of certain motivational interests. For example, one might want to say that a *consequence* of living in smaller communities is the development of a heightened interest in affiliation and yet there remains one possibility that such an interpretation is in error. Motivation can play a large role in *determining* the kind of community people live in, whether they are willing to remain in a city or in a rural area, whether they migrate to suburban areas, and the like. For example, men with high affiliation motivation may prefer small towns where intimacy is easier.

### Broken Home Background and Death of a Spouse

In the foregoing sections we have suggested that many of the differences in motivation scores obtained can be as plausibly interpreted as situational differences as they can as personality differences. The "correctness" of either interpretation, or the relative importance of each interpretation should be resolved in further study. There is evidence that the resolution of this problem will not be a general one; one or the other interpretation will not be uniformly appropriate, but both will be involved. This evidence comes from examining two other special groupings—broken home background, and death of a spouse.

All respondents were asked whether or not they had lived with both of their parents

before the age of 16. Through further probing we were able to group respondents into three categories:

1. Coming from intact homes.
2. One or both parents died before the respondent was 16.
3. Parents were divorced or separated before respondent was 16.

These experiences in early childhood are likely to be profound catalysts for inducing basic personality patterns. We feel that differences found in adults with these three kinds of early backgrounds might reflect early personality integration of motivation. In these comparisons (Table 21), parental divorce seems to have had effect on achievement motivation—lowering it for males and raising it for females. Death in the family lowers affiliation motivation in males, but not in females, and either death or divorce raises power motivation in males and lowers it in females. We have found a sizable relationship between the age of the respondent and experiences of family disruption by divorce. Proportionally more younger than

older respondents make up the group from divorce backgrounds. Even with age controlled, however, most of the differences apparent in Table 21 remain; for the most part, the differences cited occur at each age group considered.

Explanations of these differences would rest, we argue, on considering what implications these early life experiences have on motive development. For example, we might suggest that divorce of one's parents can have a lasting effect on a child's achievement motivation because the disruption creates significant early experiences and attitudes towards the competence of both the father and the mother. Usually children remain with the mother when there is a divorce. Resentment or criticism of the father must be inevitably transmitted to the child. This resentment added to the mere absence of an intimate contact with the father can have opposite effects on the achievement motivation of boys and girls. A boy, having lost his masculine model for achievement, may become highly involved in avoiding failure. In so doing, his achievement motivation, his positive motivations for success, become weakened. On the other hand, girls living with a divorced mother have a readily available model for achievement identification. Resentment of the father can reinforce a need for feminine independence and self-reliance in a masculine world. The fact that her mother is apparently self-sufficient further enhances an image of the achievement orientation of women. This is overgeneralized speculation about the dynamics of divorce and achievement motivation. Nevertheless, it is an example of how we would treat differences in motives found in family background comparisons. Having obtained differences in motivation scores for these groups we can proceed with an interpretation of differences by assuming that the differences reflect enduring *personality* changes brought on by family disruptions in early childhood. What is perhaps most important is that these potential differences in motivation are attributable to dispositional differences. Therefore, the motivation score *can* reflect personality differences.

TABLE 21

BROKEN HOME BACKGROUND RELATED TO  
N ACHIEVEMENT, N AFFILIATION, AND N POWER

Motive and Home Background	Percentage of High Scores*	
	Men	Women
n Achievement		
Intact home	54	48
Parent(s) died	45	42
Parents divorced or separated	38	60
n Affiliation		
Intact home	49	49
Parent(s) died	39	52
Parents divorced or separated	46	51
n Power		
Intact home	47	51
Parent(s) died	61	45
Parents divorced or separated	62	44

\* Does not include Ss whose home background was not ascertained or Ss whose home background was not codable into these categories. Intact home Ns, Men = 471, Women = 585; Parent(s) died Ns, Men = 80, Women = 125; Parents divorced Ns, Men = 24, Women = 45.

The last table we are going to consider is one which compares two groups of men and women—those presently married and those whose spouses have died (Table 22). We present this table only for older (50 and above) men and women because the number of widows and widowers under 50 is relatively low. Differences without this age control might be attributable to age rather than the loss of a spouse. There are interesting differences in the table. For all motives, except power motivation in men, the loss of a spouse yields a decrease in score which is primarily a situational change. The general conclusion might be that losing one's wife or husband in later life for most people has the effect of decreasing their motivational concerns—a conclusion not at all out of line with current thinking about problems in old age. At any rate we have evidence that a life situational factor in later life can have a strong effect on motivation scores.

#### SUMMARY AND CONCLUSIONS

We have reviewed some of the methodological and interpretive problems attending the use of thematic apperceptive measures of motivation in a nationwide interview study. In light of this exploration, it is possible to offer both evaluations of the difficulties encountered and a prognosis concerning these measures in substantive research within the survey setting.

*Selection of Pictures.* The selection of separate sets of six pictures for men and for women was guided by previous research concerning the effect of picture cues on thematic apperception. The percentage of stories elicited by these pictures which contained imagery related to achievement, affiliation, and power was lower than anticipated from pretests but sufficient, we believe, to allow for valid assessment of individual and group differences in the strength of the three motives.

We have no independent basis for evaluating the extent to which the objective was attained of providing a fair opportunity for persons in all strata of American society to express their motives. Only intensive analysis of the responses to particular pictures by particular segments of the population (reserved for a later paper) will provide the kind of evidence needed for further evaluation of the adequacy of the pictures that were employed and clarification of the issues to be faced in subsequent attempts of this sort.

*Interviewer Effects.* In light of the known sensitivity of thematic apperception to conditions of administration, we anticipated and found that despite our efforts to standardize the interviewing procedure, the interviewers did (apparently) contribute significantly to the variance of some (but not all) of the motivation scores. However, we were gratified to find that our estimates

TABLE 22  
MARITAL STATUS RELATED TO n ACHIEVEMENT, n AFFILIATION, AND n POWER  
(Age 50 and Above Only)

Motive	Percentage of High Scores			
	Men		Women	
	Married (N=169)	Widowed (N=27)	Married (N=122)	Widowed (N=93)
n Achievement	47	37	47	43
n Affiliation	50	41	43	39
n Power	48	63	47	43

of interviewer effect—maximal estimates as a consequence of possible confounding of interviewer and respondent effects in survey data—were no higher than in fact they were. That interviewer effects on motivation scores were less than anticipated can be attributed to the correction employed to remove the relationship between raw motivation scores and length of protocol. Interviewers differed considerably in the average length of the imaginative protocols they obtained. Correction of motivation scores for this factor effectively eliminated much of an otherwise substantial interviewer effect.

As a consequence of bias attributable to interviewers still remaining in some of the motivation scores, we anticipate some loss in the sensitivity of the measures (i.e., an increase in error variance) in later analyses of substantive relationships, e.g., relationships of motivation to demographic variables.

*Reliability of Coding.* The reliability of coding attained by a team of nine relatively novice student coders, while less than that normally attained with this method of content analysis in small-scale experimental studies of college students, was nevertheless encouragingly high. The range of coding reliability coefficients for the motivation scores assigned to individuals was .72 to .91 with a median of .77. All of these estimates of coding reliability are within the range of reliabilities reported in earlier published studies. However, in comparison with the median standard of .89 attained in experimental studies, in this study with new pictures and other novel problems there is a loss in reliability of coding and hence some increase in the error of measurement. We expect, therefore, an additional loss in the sensitivity of the measures in later analyses of substantive relationships.

*Inadequate Protocols.* A substantial number of individuals produced protocols that were judged inadequate for further analysis. The over-all loss of data amounted to 17% among men and 14% among women. In both sexes, the loss was heavy and serious, as might have been anticipated,

only among the least educated groups. Hence, the average motivation scores of some subgroups in the population, specifically those in which level of education is very low, must be considered biased estimates—derived only from the most verbal members of those groups.

*Verbal Fluency.* One problem of considerable importance, the relationship between motivation scores and length of imaginative protocol, which is described by correlations of .20 to .28 in this national sample, seems to have been quite adequately overcome. The correction devised reduces these correlations to the range from  $-.04$  to  $.05$  and effectively removes the possibility that substantive relationships, like the significant positive relationship between Achievement and education of respondent, can be considered spurious because of a dependency of motivation scores on verbal fluency. As mentioned earlier, this correction also produced a very substantial reduction in the extent of variability of motivation scores attributable to interviewer differences.

*Interpretation of Motivation Scores.* Differences in motivation scores in various subgroups of the population can be variously interpreted as reflecting differences in: personality, life situation, and the significance of the pictures for various groups. The relative importance of each of these interpretations may be evaluated only by further inquiry into the kinds of variables related to these scores within different subgroups. A particular network of interrelationships may be suggestive of which of these views may be supported. It might very well be that the interpretation of the scores will vary depending on which social group is being investigated. Or, one conclusion that seems likely is that for each social group the motivation score can be partially interpreted as personality assessment and partially as an assessment of reactions to ongoing life situations. There is evidence in the survey data that each of these interpretations may be operating in the scores. Motivation scores reflect differences both in early life experiences (broken home back-

ground) and later life experiences (death of a spouse).

These alternative considerations of the conceptual meaning of the motivation scores precludes the consideration of their relationships with demographic variables as unequivocal norms describing subgroup differences in personality. The results have been presented here with the hope that they can serve as background for exploring the significance of motivation differences in society.

If methodological precautions like those outlined here are seriously considered, the use of thematic apperceptive measures of motivation (and other sensitive techniques of personality assessment) can be expected to make marked contributions in the study

of important substantive problems in survey research. Careful inquiries into the relationship of these measures to other psychological variables and to demographic indices in this survey, and in subsequent national studies, should greatly enhance our understanding of the social origins and consequences of motivation. At the same time, survey studies will provide insights that are needed to refine the thematic apperceptive method itself. The use of thematic apperception to assess motivation in survey studies of national character will mean that factual evidence concerning personality configurations within American society can be integrated with the fund of factual evidence concerning the dynamics of motivation already derived from experimental and clinical use of these same measures.



## REFERENCES

- ANDERSON, R. L., & BANCROFT, T. A. *Statistical theory in research*. New York: McGraw-Hill, 1952.
- ANGELINA, A. L. Un novo método par avaliar a motivação humana. Unpublished doctoral dissertation, Univer. São Paulo, Brazil, 1955.
- ATKINSON, J. W. Studies in projective measurement of achievement motivation. Unpublished doctoral dissertation, Univer. Mich., 1950.
- ATKINSON, J. W. (Ed.) *Motives in fantasy, action and society: A method of assessment and study*. Princeton, N. J.: Van Nostrand, 1958.
- BIRNEY, R. C. Experimenter effect on the achievement motive. Unpublished manuscript, Amherst Coll., 1956.
- BIRNEY, R. C. Thematic content and the cue characteristics of pictures. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. Princeton, N. J.: Van Nostrand, 1958. Pp. 630-643.
- CAMPBELL, D. T. The indirect assessment of social attitudes. *Psychol. Bull.*, 1950, **47**, 15-38.
- CHILD, I., STORM, T., & VEROFF, J. Achievement themes in folk tales related to socialization practice. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. Princeton, N. J.: Van Nostrand, 1958. Pp. 479-492.
- CHRISTIE, R. Acquiescence and education. Unpublished manuscript, Columbia Univer., 1958.
- COOK, R. A. Identification and ego defensiveness in thematic apperception. *J. proj. Tech.*, 1953, **17**, 312-319.
- DAVENPORT, J. W. The projective expression of the achievement motive in relation to career orientation and social satisfaction in home economics freshmen. Unpublished master's thesis, Univer. Mass., 1953.
- DOUVAN, ELIZABETH. Social status and success strivings. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. Princeton, N. J.: Van Nostrand, 1958. Pp. 509-517.
- DOUVAN, ELIZABETH, & WALKER, A. M. The sense of effectiveness in public affairs. *Psychol. Monogr.*, 1956, **70**(22, Whole No. 429).
- FELD, SHEILA, & SMITH, C. P. An evaluation of the method of content analysis. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. Princeton, N. J.: Van Nostrand, 1958. Pp. 234-241.
- FELD, W. F. The effects of thematic apperception upon certain experimentally aroused needs. Unpublished doctoral dissertation, Univer. Md., 1951.
- GURIN, G., VEROFF, J., & FELD, SHEILA. *Americans view their mental health: A nationwide interview survey*. New York: Basic Books, 1960.
- HABER, R. N., & ALPERT, R. The role of situation and picture cues in projective measurement of the achievement motive. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. Princeton, N. J.: Van Nostrand, 1958. Pp. 644-663.
- JACOBS, B. A method for investigating the cue characteristics of pictures. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. Princeton, N. J.: Van Nostrand, 1958. Pp. 617-629.
- KISH, L. Selection of the sample. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences*. New York: Dryden, 1953. Pp. 175-240.
- KORCHIN, S. J., MITCHELL, H. E., & MELTZOFF, J. A critical evaluation of the Thompson Thematic Apperception Test. *J. proj. Tech.*, 1950, **14**, 445-452.
- LANSING, J., & HEYNS, R. Need affiliation and frequency of four types of communication. *J. abnorm. soc. Psychol.*, 1959, **58**, 365-372.
- LIGHT, B. H. A further test of the Thompson TAT rationale. *J. abnorm. soc. Psychol.*, 1955, **51**, 148-150.
- LINDZEY, G., & HEINEMANN, S. H. Thematic Apperception Test: Individual and group administration. *J. Pers.*, 1955, **24**, 34-55.
- MACCOBY, ELEANOR, E., & MACCOBY, N. The interview: A tool of social science. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. 1. Cambridge: Addison-Wesley, 1954. Pp. 449-487.
- MCCLELLAND, D. C. Risk taking in children with high and low need for achievement. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. Princeton, N. J.: Van Nostrand, 1958. Pp. 306-321.
- MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
- MCCLELLAND, D. C., & FRIEDMAN, G. A. A cross-cultural study of the relationship between child-training practices and achievement motivation appearing in folk tales. In G. E. Swanson, T. M. Newcomb, & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1952. Pp. 243-249.
- MORGAN, J. N., SNIDER, M., & SOBEL, M. G. *Highlights from a study on lump sum redemption settlements and rehabilitation*. Ann Arbor, Mich.: Survey Research Center, 1958.
- MORRISON, H. W. Validity and behavioral correlates of female need for achievement. Unpublished master's thesis, Wesleyan Univer., 1954.

- MURRAY, H. A. *Thematic Apperception Test manual*. Cambridge: Harvard Univer. Press, 1953.
- REITMAN, W. R., & ATKINSON, J. W. Some methodological problems in the use of thematic apperceptive measures of human motives. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. Princeton, N. J.: Van Nostrand, 1958. Pp. 664-683.
- RICCIUTI, H. N. *The prediction of academic grades with a projective test of achievement motivation: I. Initial validation studies*. Princeton, N. J.: Educational Testing Service, 1954.
- RICCIUTI, H. N., & SADACCA, R. *The prediction of academic grades with a projective test of achievement motivation: II. Cross validation at the high school level*. Princeton, N. J.: Educational Testing Service, 1955.
- RIESS, B. F., SCHWARTZ, E. K., & COTTINGHAM, ALICE. An experimental critique of assumptions underlying the Negro version of the TAT. *J. abnorm. soc. Psychol.*, 1950, **45**, 700-709.
- ROSEN, B. The achievement syndrome: A psychocultural dimension of social stratification. *Amer. sociol. Rev.*, 1956, **21**, 203-211. (Also in J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. Princeton, N. J.: Van Nostrand, 1958. Pp. 495-509.)
- SANFORD, F. H. The use of a projective device in attitude surveying. *Publ. opin. Quart.*, 1950, **14**, 697-709.
- SANFORD, F. H., & ROSENSTOCK, J. Projective techniques on the doorstep. *J. abnorm. soc. Psychol.*, 1952, **47**, 3-16.
- SCHWARTZ, E. K., RIESS, B. F., & COTTINGHAM, ALICE. Further critical evaluation of the Negro version of the TAT. *J. proj. Tech.*, 1951, **15**, 394-400.
- SCOTT, W. A. The avoidance of threatening material in imaginative behavior. *J. abnorm. soc. Psychol.*, 1956, **52**, 338-346.
- SNEDECOR, G. W. *Statistical methods*. Ames, Iowa: Iowa State College Press, 1946.
- TOMPKINS, S. S., & MINER, J. B. *The Tompkins-Horn Picture Arrangement Test*. New York: Springer, 1958.
- VEROFF, J., WILCOX, SUE, & ATKINSON, J. W. The achievement motive in high school and college age women. *J. abnorm. soc. Psychol.*, 1953, **48**, 108-119.
- VOGEL, MARGARET. An investigation of the affiliation motive in college age women using low cue strength pictures. Unpublished honors' thesis, Univer. Mich., 1954.
- WALKER, E. L., & ATKINSON, J. W. The expression of fear-related motivation in thematic apperception as a function of proximity to an atomic explosion. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. Princeton, N. J.: Van Nostrand, 1958. Pp. 143-159.

(Received January 12, 1960)

## APPENDIX

## APPROXIMATE SAMPLING ERROR DIFFERENCES FOR PERCENTAGES FROM 35% TO 65%

N	700	55	300	200	100	75	50	25
700	5.4-7.6	5.9-8.3	6.9-9.7	8.0-11.2	10.7-14.4	12.1-15.9	14.6-18.9	20.3-25.8
500		6.3-8.8	7.2-10.1	8.4-11.8	11.0-14.8	12.3-16.2	14.8-19.2	20.5-26.1
300			8.2-11.5	9.1-12.7	11.5-15.6	12.9-17.0	15.4-19.7	20.8-26.5
200				10.0-14.0	12.2-16.7	13.4-17.9	15.8-20.7	21.2-27.2
100					14.1-19.0	15.2-20.1	17.3-22.4	22.4-28.6
75						16.2-21.2	18.2-23.5	23.1-29.4
50							20.0-25.7	24.5-31.1
25								28.3-35.8

Note.—The values shown are the differences required for significance (two standard errors) in comparisons of percentages derived from two different subgroups of the current survey. Two values—low and high—are given for each cell. The lower estimates are based on simple random samples. The higher values are based on the computation of individual sampling errors carried out on the current study data, and allow for the departure from simple random sampling in the survey design such as stratification and clustering. The sampling error does not measure the total error involved in specific survey estimates since it does not include non-response and reporting errors.

ve  
oc.

M,  
ro  
5,

ce-  
oc.

es,

as-  
k:

he  
ge  
8,

a-  
ow  
rs'

x-  
tic  
an  
,),  
ce-  
59.

---

---

8  
1  
5  
2  
6  
4  
1  
8

ges  
tes  
on  
and  
on-

S

0

